

Zur Schweizer Stichprobe PISA 2015 Eine externe Validierung zentraler Stichprobenmerkmale

Martin Verner, Universität Zürich

Andrea B. Erzinger, Universität Bern

Ursina Fässler, Pädagogische Hochschule St.Gallen

Internationale Schulleistungsstudien stellen nicht individuelle Leistungsbeurteilungen, sondern Aussagen über Populationen in den Vordergrund. Aus praktischen Gründen bietet sich dabei die Ziehung von Stichproben an. Valide Aussagen über Populationen auf Basis von Stichproben setzen methodisch korrekte Ziehungs- und Gewichtungungsverfahren voraus. Dieser Beitrag zeigt am Beispiel von PISA 2015 Qualitätsmerkmale wissenschaftlicher Zufallsstichproben auf. Dabei wird das in der Schweiz angewendete Stichprobenziehungsverfahren beschrieben. Darüber hinaus wird überprüft, inwieweit die Schweizer Stichprobe PISA 2015 die Zielpopulation in Bezug auf Schulprogramme, Migrationshintergrund und Anderssprachigkeit adäquat abbildet. Die Analysen stützen die Gültigkeit der Schweizer Stichprobe PISA 2015.

Einleitung

Im Rahmen der educational accountability stellt die evidenzbasierte Forschung ein Kennzeichen aktueller Bildungssysteme dar. Dabei wird davon ausgegangen, dass die Leistungsfähigkeit dieser Systeme an den Leistungen der Schülerinnen und Schüler gemessen werden kann (Ryan & Shepard, 2008). Auf internationaler Ebene wurde diesem Ansatz mit verschiedenen Schulleistungsstudien, an denen sich in den letzten Jahrzehnten zahlreiche Länder beteiligten, Rechnung getragen (Rutkowski, von Davier, & Rutkowski, 2014). In diesem Zusammenhang nimmt die Schweiz seit dem Jahr 2000 an PISA (Programme for International Student Assessment), einer von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) organisierten, internationalen Schulleistungstudie, teil. PISA wird alle drei Jahre in einem Grossteil der OECD-Mitgliedstaaten und einer wachsenden Anzahl von Partnerstaaten durchgeführt. Die Studie erfasst alltags- und berufsrelevante Kompetenzen von 15-Jährigen in den Domänen Mathematik, Lesen und Naturwissenschaften. Weiter werden diverse, bildungspolitisch relevante Hintergrundinformationen

zu Schulen und zur Schülerschaft erhoben. An der sechsten PISA-Erhebung im Jahr 2015 nahmen weltweit mehr als eine halbe Million Jugendliche aus 72 Ländern teil.

Die internationalen Ergebnisse zu PISA 2015 wurden im Dezember 2016 (OECD, 2016a, 2016b) veröffentlicht. Vor allem der im Vergleich zu PISA 2012 um 17 Punkte niedrigere Schweizer Mittelwert in der Leseleistung (2012: 509; 2015: 492) hat dazu geführt, die Gültigkeit der Ergebnisse zu hinterfragen. Dabei wurde einerseits die Frage nach der Vergleichbarkeit der Messungen in PISA 2012 und PISA 2015 aufgrund der Umstellung auf computerbasiertes Testen aufgeworfen. Andererseits wurden die Verringerung der Stichprobengrösse und der höhere Anteil anderssprachiger¹ Schülerinnen und Schüler bei PISA 2015 im Vergleich zu PISA 2012 als Argumente ins Feld geführt, um die Qualität der für PISA 2015 gezogenen Stichprobe in Frage zu stellen (Dachverband Lehrerinnen und Lehrer Schweiz, 2016; Schweizerische Konferenz der kantonalen Erziehungsdirektoren [EDK], 2016; Neue Zürcher Zeitung, 2016; Ostschweiz am Sonntag, 2016; Tagblatt, 2016).

Vor diesem Hintergrund werden im vorliegenden Beitrag die Qualitätsmerkmale von Stichproben in Schulleistungsstudien sowie deren Voraussetzungen herausgearbeitet. Zudem wird die Bedeutung dieser Qualitätsmerkmale für das Treffen von validen Aussagen über die untersuchte Population am Beispiel von PISA 2015 aufgezeigt.² Die Qualität der Stichprobe PISA 2015 wird in einem weiteren Schritt durch eine externe Validierung der spezifischen Repräsentativität (Bortz, 1999) – in Bezug auf Merkmale, die mit den Leistungen der Schülerinnen und Schüler im Zusammenhang stehen, wie die Häufigkeit besuchter Schulstufen und Schulprogramme, die Anteile von Jugendlichen mit Migrationshintergrund und die Quoten anderssprachiger Schülerinnen und Schüler – geprüft. Diese externe Validierung erfolgte mit Daten des Bundesamts für Statistik (BFS). Abschliessend werden die Vergleiche diskutiert.

Qualitätsmerkmale von Stichproben in Schulleistungsstudien am Beispiel von PISA 2015

Ausgangspunkt jeder Stichprobenziehung stellt die Zielpopulation dar. Die PISA-Zielpopulation jedes teilnehmenden Landes besteht aus 15-Jährigen³, die eine ausbildende Institution ab dem siebten Schuljahr besuchen. In der Schweiz setzte sich diese Population im Schuljahr 2015/16 gemäss der Liste der für PISA wählbaren Schulen, generiert auf der Grundlage der Schweizerischen Lernendenstatistik⁴ durch das BFS, aus 83'542 Schülerinnen und Schülern zusammen. Dabei ist anzumerken, dass nicht sämtliche Jugendlichen, die diesem Einschlusskriterium entsprechen, noch in der obligatorischen Schule unterrichtet werden. So werden im Rahmen der PISA-Erhebungen in der Schweiz über die Schulen auf Sekundarstufe I hinaus auch gymnasiale Maturitätsschulen, Fachmittel-

schulen (FMS) oder Institutionen, die Berufsbildungsgänge (Vorbereitung auf die Berufsmaturität) oder berufliche Grundbildungen (Lehre, Berufsfachschule, Attestausbildung) anbieten, miteinbezogen.

Jedes an PISA teilnehmende Land kann bei Bedarf und einer Genehmigung des internationalen Konsortiums maximal fünf Prozent der Zielpopulation ausschliessen (OECD, 2017). So kommt es in einzelnen Ländern vor, dass entlegene Regionen oder Sprachminderheiten nicht berücksichtigt werden. In der Schweiz werden Sonderschulen sowie Bildungsinstitutionen, die nicht in einer der Schweizer Landessprachen unterrichten, ausgeschlossen (Konsortium PISA.ch, 2018). Darüber hinaus können die zur PISA-Teilnahme aufgebotenen Schülerinnen und Schüler von der Erhebung befreit werden, wenn gemäss Information durch die entsprechende Schulleitung bestimmte Kriterien zutreffen. So werden kognitiv oder funktional beeinträchtigte Jugendliche ausgeschlossen, für die eine Teilnahme an standardisierten Erhebungen durch die schulverantwortliche Person als nicht zumutbar eingeschätzt wird. Nicht untersucht werden zudem Jugendliche, die weniger als ein Jahr in der entsprechenden Testsprache unterrichtet werden und deshalb über mangelnde Kenntnisse in dieser Sprache verfügen (OECD, 2017; Konsortium PISA.ch, 2018). Aufgrund dieser Ausschlüsse umfasst der Umfang der tatsächlichen PISA-Population in der Schweiz knapp 80'000 15-Jährige (Konsortium PISA.ch, 2018).

Präzision der Schätzung eines Populationsparameters als Qualitätsmerkmal einer gezogenen Stichprobe am Beispiel von PISA 2015

Da es aus ökonomischen sowie organisatorischen Gründen nicht möglich ist, dass die gesamte Population der 15-jährigen Schülerinnen und Schüler in der Schweiz an einer PISA-Erhebung teilnimmt, werden im Vorfeld der jeweiligen PISA-Erhebungen Stichproben gezogen (siehe auch Rust, 2014; Rutkowski, Gonzalez, Joncas, & von Davier, 2010). Gemäss Bortz (1999, S. 86) stellt eine Stichprobe eine «Teilmenge aller Untersuchungseinheiten dar, die die untersuchungsrelevanten Eigenschaften der Grundgesamtheit möglichst genau abbilden soll» (siehe auch Kruskal & Mosteller, 1979a, 1979b, 1980; Schnell, Hill, & Esser, 1999). Die auf PISA beruhenden Schätzungen zu diesen Eigenschaften stimmen nicht immer präzise mit den tatsächlichen Populationswerten überein, da die PISA-Stichproben auf einer Zufallsauswahl beruhen. Nebst zahlreichen statistischen Vorzügen liegen die Vorteile dieser Zufallskomponente hauptsächlich darin, dass die Präzision der Schätzungen bzw. die Stichprobenfehler quantifiziert und gesteuert werden können (von der Lippe & Kladroba, 2002). Die Steuerung erfolgt – neben dem Stichprobenziehungsverfahren, welches weiter unten beschrieben wird – über den Stichprobenumfang, der im Rahmen von PISA in erster Linie von der OECD festgelegt wird. Diese setzt aktuell eine Mindestanzahl von 150 einbezogenen Schulen und 4'500 getesteten Jugendlichen pro teilnehmendem Land voraus (OECD, 2017).

Darüber hinaus haben die einzelnen Länder je nach individuellen nationalen Forschungsinteressen die Möglichkeit, Zusatzstichproben zu untersuchen. So wurden in der Schweiz in den ersten fünf PISA-Durchgängen zwischen 2000 und 2012 Zusatzstichproben gezogen, die Leistungsvergleiche zwischen Kantonen auf Basis der Schülerschaft des letzten obligatorischen Schuljahres erlaubten. Seit PISA 2015 wurde aufgrund einer bildungspolitischen Entscheidung auf diese Zusatzstichproben verzichtet und eine Stichprobe ausschliesslich für die 15-jährigen Schülerinnen und Schüler gezogen, was ungefähr einer Halbierung des Stichprobenumfangs im Vergleich zur Erhebung PISA 2012 entsprach (EDK, 2009, 2012).

Da die vergleichsweise komplexeren Ziehungsverfahren⁵ vergangener PISA-Erhebungen die diesbezüglichen Vorteile einer grösseren Stichprobe aufwogen, führte die Reduktion der Anzahl Studienteilnehmenden im Hinblick auf PISA 2015 und die damit verbundene Vereinfachung des Ziehungsdesigns nicht zu einer Einbusse in der Präzision geschätzter, nationaler Leistungsmittelwerte. So weisen die in Tabelle 1 dargestellten Standardfehler (die sich aus Stichproben- und Messfehler zusammensetzen) auf keine Differenzen in der Schätzpräzision zwischen den Erhebungsjahren 2006⁶ und 2015 hin. Die äusserst umfangreichen Stichproben der früheren Erhebungen zahlten sich demnach vor allem bei Auswertungen von kleineren Gruppen aus (z.B. Kantons- oder Schultypenvergleiche). Der deutlich kleinere Stichprobenumfang bei PISA 2015 ist jedoch für statistische Analysen und internationale Vergleiche ausreichend.

Tabelle 1. Stichprobenumfänge sowie Standardfehler der schweizerischen Gesamtmittelwerte für die drei bei PISA erhobenen Domänen, getrennt für die PISA-Erhebungsjahre 2006 bis 2015

PISA	N	Standardfehler des Mittelwerts		
		Lesen	Mathematik	Naturwissenschaften
2006	12'192	3.1	3.2	3.2
2009	11'812	2.4	3.3	2.8
2012	11'229	2.6	3.0	2.7
2015	5'860	3.0	2.9	2.9

Anmerkung. Die Stichprobengrösse (N) beinhaltet effektiv teilnehmende Schülerinnen und Schüler. Für die Erhebung aufgebotene Schülerinnen und Schüler, die aus verschiedenen Gründen (Krankheit, Wegzug, Ausschluss) abwesend waren, werden hier nicht einberechnet. Die Zahlen beziehen sich auf den internationalen Datensatz und dementsprechend auf die Grundgesamtheit der 15-jährigen Schülerinnen und Schüler. Die dargestellten Standardfehler setzen sich aus Stichproben- und Messfehler zusammen.

Es lässt sich demnach festhalten, dass die Präzision der Schätzung eines Populationsparameters nicht nur vom Stichprobenumfang, sondern auch vom Stichprobenziehungsverfahren abhängig ist (Bortz, 1999) und die Reduktion des Stichprobenumfangs im Hinblick auf PISA 2015 keine Qualitätseinbussen in internationalen Leistungsvergleichen zur Folge hatte. Daher wird in den

folgenden Abschnitten das im Rahmen von PISA 2015 eingesetzte Stichprobenziehungsverfahren genauer betrachtet.

Stichprobenziehungsverfahren als Qualitätsmerkmal einer gezogenen Stichprobe am Beispiel von PISA 2015

Bei PISA wird in den meisten teilnehmenden Ländern ein zweistufiges Stichprobenziehungsverfahren mit Stratifizierung eingesetzt. Dabei werden in einem ersten Schritt die teilnehmenden Schulen gezogen, bevor 15-Jährige – innerhalb der gezogenen Schulen – systematisch für die PISA-Erhebung bestimmt werden. Um die Effizienz – im Sinne von präzisen Schätzern – des Stichprobenziehungsverfahrens zu erhöhen und sicherzustellen, dass untersuchungsrelevante Charakteristiken der Population adäquat in der Stichprobe vertreten sind, wird auf beiden Stufen des Stichprobenziehungsverfahrens die Population nach bestimmten, meist mit der Schulleistung in Zusammenhang stehenden Merkmalen geschichtet (stratifiziert; für weiterführende Informationen siehe OECD, 2009, 2017).

Im Rahmen von PISA 2015 wurde auf der Ebene der Schulen sowohl auf explizite als auch auf implizite Stratifizierungsvariablen zurückgegriffen, wie dies bereits bei den Erhebungen in den Vorjahren der Fall war (OECD, 2017). Die explizite Stratifizierung beinhaltet die Bildung von Schichten (Strata), die unabhängig voneinander im Verlauf des Stichprobenziehungsprozesses behandelt werden. Für die Ziehung der Schulstichprobe PISA 2015 in der Schweiz wurden die folgenden Schulmerkmale als explizite Stratifizierungsgrundlage verwendet:

- Sprachregion mit drei Stufen: Deutsch, Französisch, Italienisch.
- ISCED-Klassifikation (International Standard Classification of Education) mit drei Stufen: Schulen der Sekundarstufe I (z.B. Real-, Ober- oder Orientierungsschulen), Schulen der Sekundarstufe II (z.B. Berufs- oder Fachmittelschulen), gemischte Schulen (z.B. Kantonsschulen).
- Trägerschaft mit drei Stufen: Öffentliche Schulen, private Schulen mit einem Subventionierungsgrad von über 50 Prozent und private Schulen mit einem Subventionierungsgrad von unter 50 Prozent.

Dieses 3x3x3-Design resultierte in 27 unterschiedlichen Strata. Innerhalb dieser Strata wurden die Schulen beruhend auf der Schultypkombination, nach Kanton und schliesslich nach Schulgrösse implizit stratifiziert (Rust, 2014, S. 129), was in einer im Voraus definierten, festen Abfolge resultierte.⁷ Die Schultypkombinationen ergaben sich aus den innerhalb einer Schule angebotenen Programmen, wobei hier die Kategorisierung des BfS herangezogen wurde (Tabelle 2). Entsprechend dem generell bei PISA eingesetzten probability proportional to size-Verfahren (PPS) (OECD, 2017; Rust, 2014, S. 130) war die Auswahlwahrscheinlichkeit für Schulen mit einer umfangreichen Schülerschaft höher als für vergleichsweise kleine Schulen.

Nachdem die an PISA 2015 teilnehmenden Schulen bestimmt waren, wurden pro Schule 35 Jugendliche⁸ gezogen. Dazu wurden vollständige Listen sämtlicher unterrichteter, 15-jähriger Schülerinnen und Schüler nach Schuljahr, Geschlecht (OECD, 2017) und Klassenzugehörigkeit implizit stratifiziert. Anschliessend wurde eine systematische Zufallsstichprobe gezogen: Innerhalb jeder gezogenen Schule hatten alle Individuen dieselbe Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden. Die kombinierte Auswahlwahrscheinlichkeit (Wahrscheinlichkeiten auf Schul- und Individualebene multipliziert) war aufgrund der PPS-Stichprobe für alle Schülerinnen und Schüler eines expliziten Schulstratums annäherungsweise identisch.⁹ Auf Schülerebene wurde auf eine explizite Stratifizierung verzichtet.

Gewichtung und Stichprobenvarianz bei der Stichprobe am Beispiel von PISA 2015

Wie im letzten Abschnitt erläutert, resultiert das bei PISA verwendete Stichprobenziehungsverfahren in einer zufälligen Selektion von Schulen sowie Schülerinnen und Schülern, deren Auswahlwahrscheinlichkeiten bekannt sind. Diese Wahrscheinlichkeiten dienen als Basis zur Berechnung von Stichprobengewichten, die für das Treffen von validen Aussagen über die untersuchte Population von zentraler Bedeutung sind (Meinck, 2015). Die Stichprobengewichte berücksichtigen die teilweise unterschiedlichen Auswahlwahrscheinlichkeiten der teilnehmenden Schülerinnen und Schüler und widerspiegeln somit die Anzahl nichtgezogener 15-Jähriger, die durch die einzelnen Studienteilnehmenden vertreten werden. Die Stichprobengewichte setzen sich aus den reziproken Werten der Auswahlwahrscheinlichkeiten für Schulen sowie den Auswahlwahrscheinlichkeiten von Schülerinnen und Schülern (innerhalb der Schulen) zusammen. Dieses Produkt wird zusätzlich mit diversen Korrekturfaktoren multipliziert, mithilfe derer verweigernde Schulen sowie abwesende oder verweigernde Jugendliche mitberücksichtigt werden (Non-Response-Korrektur; OECD, 2017).

Auch bei der Schätzung des Stichprobenfehlers nimmt die Gewichtung eine zentrale Rolle ein. Aufgrund der Unsicherheit, die durch die zufällige Ziehung von Schulen sowie 15-Jährigen entsteht, werden PISA-Ergebnisse stets mit einem auf dem Stichprobenfehler beruhenden Konfidenzintervall berichtet (OECD, 2009). Der Ansatz, der zur Berechnung des Stichprobenfehlers bei PISA-Analysen herangezogen wird, ist unter dem Begriff *Balanced Repeated Replication* (BRR) (Rust, 1985; Rust & Rao, 1996; Wolter, 2007) bekannt, wobei auf Fays Variante der BRR-Methode zurückgegriffen wird (Judkins, 1990). Bei diesem mit der Jackknife-Methode (Rust, 2014) verwandten Replikationsverfahren werden teilnehmende Schulen beruhend auf den Stratifizierungsvariablen gepaart. Innerhalb jedes Paares wird anschliessend eine Einheit stärker und eine Einheit schwächer gewichtet, wobei dieses Vorgehen wiederholt wird, bis 80 unterschiedliche Kombinationen (Replikationsgewichte) aus über- und

untergewichteten Schulen entstehen. Diese 80 Kombinationen entsprechen 80 neuen Varianten der Stichprobe, die sich durch eine veränderte Gewichtung der gezogenen Schülerinnen und Schüler von der anfänglichen Stichprobe unterscheiden. Der Stichprobenfehler eines Leistungsmittelwerts – oder eines beliebigen auf dem PISA-Datensatz beruhenden Schätzers – wird aus der Variabilität der entsprechenden Werte zwischen den 80 neugebildeten Varianten der Stichprobe abgeleitet.

Externe Validierung der spezifischen Repräsentativität

Unter der Güte einer Stichprobe wird im Allgemeinen das Mass dafür verstanden, inwieweit die gezogene Stichprobe ein adäquates Abbild der Population hinsichtlich der sie charakterisierenden Merkmale darstellt (siehe auch Rendtel, 1993). Für die hier im Zusammenhang mit der Stichprobe für Schulleistungsstudien berichtete Validierung wurden die drei folgenden Merkmale herangezogen:

- Häufigkeiten besuchter Schulstufen und Schulprogramme (1)
- Anteile von Jugendlichen mit Migrationshintergrund (2)
- Quoten anderssprachiger Schülerinnen und Schüler (3)

Diese Populationsmerkmale wurden ausgewählt, weil (a) sie einen Zusammenhang mit der schulischen Leistung aufweisen und demnach als charakterisierend für Schulleistungsstudien beurteilt werden (Brühwiler & Helmke, 2018; OECD, 2016a, 2016b; Rendtel & Pöttner, 1992), (b) die mit PISA 2015 geschätzten Anteile wie in der Einleitung dargestellt hinterfragt wurden und (c) die Merkmalsanteile mit externen Quellen verglichen werden können. Entsprechend wird im Folgenden geprüft, ob die Schweizer Stichprobe PISA 2015 die genannten Merkmale adäquat abbildet.

(1) Externe Validierung der Häufigkeiten von Schulstufen und Schulprogrammen in der Schweizer Stichprobe PISA 2015

Die am stärksten mit der Schulleistung zusammenhängenden Merkmale sind die Schulstufen und -programme, weshalb diese eine zentrale Rolle im Stichprobendesign einnehmen (Forschungsgemeinschaft PISA Deutschschweiz/FL, 2005). Aus diesem Grund wird in den folgenden Abschnitten überprüft, inwieweit die Häufigkeiten von Schulprogrammen und Schulstufen zwischen der gewichteten Stichprobe PISA 2015 und der Population übereinstimmen.

Methode

Die Häufigkeiten der unterrichteten Schulstufen sowie Schulprogramme in der Population lassen sich mithilfe der Statistik der Lernenden (SDL) berechnen (Bundesamt für Statistik, 2017). Diese Erhebung umfasst unter anderem

sämtliche Schülerinnen und Schüler vom Kindergarten bis zur höheren Berufsbildung und wird seit 1977 durch das BfS in Zusammenarbeit mit den Kantonen jährlich durchgeführt. Die Erhebung liefert Daten zu Beständen der Schulstufen und Schulprogramme, zur Anzahl Schulen und Klassen sowie demografische Angaben zu den Schülerinnen und Schülern. Für den Vergleich wurde beim BfS ein Auszug der SDL mit sämtlichen Schulen, die im Schuljahr 2014/15 (Durchführungsjahr PISA 2015) 15-Jährige unterrichteten, und der jeweils entsprechenden Anzahl 15-jähriger Schülerinnen und Schüler pro Schulstufe sowie Schulprogramm eingeholt. Die Anteile der besuchten Schulstufen und Programme wurden berechnet, indem die entsprechenden absoluten Häufigkeiten ins Verhältnis zur Population der 15-Jährigen (ohne Sonderschulen und «International Schools»; $N = 80'171$) gesetzt wurden.¹⁰ Da es sich bei der SDL um eine Vollerhebung handelt, werden die Zahlen aus der Population ohne Konfidenzintervall berichtet.

Die gewichteten Häufigkeiten von Schulprogrammen und Schulstufen in der Stichprobe wurden auf Basis des öffentlich zugänglichen PISA-Datensatzes berechnet. Die dazugehörigen Konfidenzintervalle wurden beruhend auf dem zuvor erwähnten Replikationsverfahren ermittelt (BRR).

Im Verlauf des Stichprobenziehungsverfahrens wurde die Schulstufe von den zuständigen Schulleitungen oder Lehrpersonen auf den Listen gezogener 15-Jähriger angegeben. Grundsätzlich wurde hier zwischen Sekundarstufe I und Sekundarstufe II differenziert, wobei Schulen auf Sekundarstufe II weiter nach gymnasialen Maturitätsschulen, Fachmittelschulen, Brückenangeboten und diversen Berufsschulen unterschieden wurden. Dieselbe Variable wurde zusätzlich im Fragebogen, den die Jugendlichen nach den Leistungstests bearbeiteten, erfragt (Konsortium PISA.ch, 2015; OECD, 2014).¹¹ Der Abgleich der Angaben beider Informationsquellen diente hierbei als Qualitätskontrolle, wobei Inkonsistenzen in weniger als fünf Prozent der Fälle vorkamen. Im Zweifelsfall wurden die Angaben der Schule als korrekt eingestuft und in den Datensatz übernommen.

Zusätzlich wurden auf Sekundarstufe I die kantonalen Schulprogramme der Schülerinnen und Schüler erhoben. Damit die Anteile der unterschiedlichen Anforderungsniveaus verglichen werden konnten, wurden die kantonalen Schulprogramme in national vergleichbare Schultypen übersetzt. Dabei wurden dieselben Kategorien wie in der SDL verwendet, in welcher zwischen Programmen mit erweiterten Anforderungen (z.B. Sekundarschulen), Grundanforderungen (z.B. Realschulen), Sonderklassen und Programmen ohne Selektion unterschieden wird (Bundesamt für Statistik, 2017). Darüber hinaus wurden Programme, die als Vorbereitung für die gymnasiale Maturität gelten, einer Kategorie mit hohen Anforderungen (z.B. Gymnasien oder Kantonsschulen) zugeordnet. Die Kategorie für Programme mit hohen Anforderungen wird offiziell nicht in der SDL geführt und wird speziell für die PISA-Stichprobenziehungsverfahren vom BfS gebildet.

Ergebnisse

Die Prozentanteile der im Schuljahr 2014/15 von 15-Jährigen besuchten Schulstufen sowie Programme sind in Tabelle 2 dargestellt. Dabei zeigte sich, dass bei PISA 2015 der Anteil der auf Sekundarstufe I unterrichteten 15-Jährigen um ca. 2 Prozentpunkte leicht überschätzt und dementsprechend der Anteil auf Sekundarstufe II leicht unterschätzt wurde. Beide Anteile lagen jedoch im Rahmen der mit PISA 2015 geschätzten 95-Prozent-Vertrauensintervalle. Bis auf die Anteile in Sonderklassen¹² unterrichteter Jugendlicher fielen sämtliche auf der Sekundarstufe I angebotenen Schulprogramme in die durch PISA 2015 geschätzten Vertrauensintervalle. Ein ebenso symmetrisches Bild zeigte sich auf der Sekundarstufe II: Einzig der Anteil der Schülerkategorie «Berufs- oder Fachmaturitäten» war höher als das entsprechende mittels PISA 2015 geschätzte Vertrauensintervall (siehe Tabelle 2).

Tabelle 2. Anteile 15-jähriger Schülerinnen und Schüler in der Schweiz für die verschiedenen, in PISA involvierten Schulprogramme auf den Sekundarstufen I und II im Schuljahr 2014/15

Schulprogramm	Anteil in Population (Prozente)	Gewichteter Anteil in Stichprobe (Prozente)	Vertrauensintervall gewichteter Stichprobenanteil (Prozente)		
<i>Sekundarstufe I</i>	71.64	73.78	71.11	-	76.45
Programm mit hohen Anforderungen*	14.05	16.13	13.90	-	18.36
Programm mit erweiterten Anforderungen	28.59	28.91	26.48	-	31.34
Programm mit Grundanforderungen	22.95	22.07	19.87	-	24.27
Programm ohne Selektion	4.20	3.43	1.22	-	5.64
Sonderklassen	1.86	1.23	0.64	-	1.82
Programm unklar		2.01			
<i>Sekundarstufe II</i>	28.26	26.21	23.54	-	28.88
Brückenangebote	3.82	3.36	1.93	-	4.79
Gymnasiale Maturitätsschulen	10.76	12.74	10.43	-	15.05
Fachmittelschulen	1.67	0.97	0.03	-	1.91
Berufs- oder Fachmaturitäten	3.90	2.23	1.49	-	2.97
Berufslehre, Berufsfachschule 3-4 Jahre	7.88	6.76	4.60	-	8.92
Berufsbildung < 3 Jahre	0.24	0.15	0.01	-	0.29

Anmerkung. Die SDL ist eine Vollerhebung, weshalb die Populationsanteile ohne Vertrauensintervall berichtet werden. Gemäss SDL besuchen 0.09% der 15-jährigen Schülerinnen und Schüler eine «andere Zusatzausbildung» auf Sekundarstufe II.

*Die Programme mit hohen Anforderungen werden in den offiziell verfügbaren Daten der SDL mit den Programmen mit erweiterten Anforderungen zusammengefasst. Das Vertrauensintervall für die Stichprobenanteile bezieht sich auf das auf dem Stichprobenfehler beruhende 95%-Konfidenzintervall. Die Stichprobenanteile für die Programme innerhalb der Sekundarstufe I wurden nacherhoben. Einzelne Schulen haben dazu keine Angaben gemacht (Programm unklar).

(2) Externe Validierung der Häufigkeiten von 15-Jährigen mit Migrationshintergrund in der Schweizer Stichprobe PISA 2015

Da die SDL zwar die Nationalität, nicht aber den Migrationsstatus erhebt, war es nicht möglich, dieses Merkmal direkt im Stichprobenziehungsverfahren zu integrieren. Deshalb wurden die mit der Stichprobe PISA 2015 geschätzten Anteile von 15-Jährigen mit Migrationshintergrund mit Daten der Strukturerhebung des BfS validiert.

Methode

Im Rahmen von PISA wird der Migrationsstatus beruhend auf Fragebogenitems berechnet, die nach dem Geburtsort der Schülerin bzw. des Schülers und der Eltern fragen. Dabei werden drei Gruppen differenziert: Zur Gruppe «Migrationshintergrund 1. Generation» werden ausschliesslich Jugendliche gerechnet, die selbst – sowie beide Elternteile – im Ausland geboren sind. In der Schweiz geborene Jugendliche, deren beide Elternteile im Ausland geboren sind, gehören zur Gruppe «Migrationshintergrund 2. Generation». Alle anderen Schülerinnen und Schüler haben gemäss der bei PISA verwendeten Definition des Migrationsstatus keinen Migrationshintergrund (OECD, 2017, S. 299). Die gewichteten Quoten von 15-Jährigen mit Migrationshintergrund wurden auf Basis des PISA-Datensatzes¹³ berechnet. Die dazugehörigen Konfidenzintervalle wurden beruhend auf dem zuvor erläuterten Replikationsverfahren (BRR) ermittelt.

Die Vergleichswerte stammen vom Auskunftsdienst der Strukturerhebung des BfS, der von der Autorenschaft gebeten wurde, eine möglichst präzise Schätzung des Anteils 15-jähriger Jugendlicher mit Migrationshintergrund – aufgrund der oben beschriebenen Einteilung – zu berechnen. Die Strukturerhebung ist ein Teil der Volkszählung und generiert Statistiken zur Bevölkerungsstruktur auf Basis von schriftlichen Befragungen. Da nur ein bestimmter Jahrgang in die Analyse integriert wurde (Jahrgang der bei PISA 2015 untersuchten Schülerinnen und Schüler: 1999), mussten trotz der jeweils umfangreichen Stichproben mehrere Erhebungsjahre kumuliert werden, um eine hinsichtlich Präzision mit PISA vergleichbare Schätzung zu berechnen: Während die aus PISA stammende Schätzung auf Jugendlichen mit Jahrgang 1999 im Schuljahr 2014/15 beruht, wurde die Schätzung der Strukturerhebung auf Basis von schriftlichen Angaben von $N = 3'523$ Jugendlichen – ebenfalls mit Jahrgang 1999 – aus den Jahren zwischen 2010 und 2014 kalkuliert.

Ergebnisse

Die Schätzungen der Anteile zum Migrationsstatus sowie die entsprechenden Vertrauensintervalle sind in Tabelle 3 dargestellt. In der Stichprobe PISA 2015 betrug der gewichtete Anteil 15-Jähriger mit Migrationshintergrund 30.33 Prozent. Die Schweizerische Strukturerhebung schätzte denselben Anteil auf 29.45 Prozent. Vergleiche getrennt nach Migrationshintergrund erster und zweiter Generation weisen auf leicht höhere Differenzen hin. So wurden die Anteile der 15-Jährigen erster Generation bei PISA 2015 niedriger und die Anteile der Jugendlichen zweiter Generation höher geschätzt als bei der Strukturerhebung. Bei Berücksichtigung der Stichprobenfehler beider Erhebungen handelt es sich hierbei insofern um nicht bedeutsame Unterschiede, als dass die Differenzen auf die Zufallskomponente der beiden Stichprobenziehungsverfahren zurückgeführt werden können. Knapp zwei Prozent der bei PISA 2015 teilnehmenden Jugendlichen beantworteten nicht genügend Fragen, um einen Migrationsstatus bestimmen zu können.

Tabelle 3. Migrationsstatus von 15-jährigen Schülerinnen und Schülern in der Schweiz getrennt für Schätzungen beruhend auf der PISA-Erhebung 2015 sowie der Schweizerischen Strukturerhebung

Migrationsstatus	Prozente	Vertrauensintervall
<i>PISA 2015</i>		
Ohne Migrationshintergrund	67.73	65.36 - 70.10
Migrationshintergrund 2. Generation	20.22	18.32 - 22.12
Migrationshintergrund 1. Generation	10.11	9.17 - 11.05
Migrationshintergrund unbekannt	1.93	
<i>Strukturerhebung 2010-2014</i>		
Ohne Migrationshintergrund	70.55	68.95 - 72.15
Migrationshintergrund 2. Generation	17.45	16.12 - 18.78
Migrationshintergrund 1. Generation	12.00	10.82 - 13.18

Anmerkung. Die Vertrauensintervalle beziehen sich auf das auf dem Stichprobenfehler beruhende 95%-Konfidenzintervall. Die mit PISA berechneten Anteile beruhen auf Fragebogenangaben; nicht alle Schülerinnen und Schüler haben die Fragen zum Geburtsort beantwortet.

(3) Externe Validierung der Anteile von anderssprachigen Schülerinnen und Schülern in der Schweizer Stichprobe von PISA 2015

Der Anteil anderssprachiger Schülerinnen und Schüler innerhalb der Schweizer Schulhäuser variiert stark. Während er in einigen Schulen verschwindend klein ist, kann er in anderen Schulen – vor allem in grösseren Städten – weit über 50 Prozent betragen. Da das PISA-Verfahren zur Ziehung der Schulstichprobe ausschliesslich die Variablen Sprachregion, Kanton, Stufe, Schulprogramm und

Schulgrösse berücksichtigte, bestand das Risiko, zufälligerweise unverhältnismässig viele Schulen mit einem sehr hohen bzw. sehr niedrigen Anteil anderssprachiger 15-Jähriger zu ziehen und dadurch die Schätzung des Populationsanteils zu verfälschen. Auf der zweiten Stufe des Stichprobenziehungsverfahrens – Schülerinnen und Schüler wurden systematisch innerhalb eines Schulhauses gezogen – ist das Risiko, dass anderssprachige Jugendliche überrepräsentiert sind, hingegen verhältnismässig gering. Deshalb wird in der Folge der mit PISA 2015 geschätzte Anteil anderssprachiger Jugendlicher mit Daten auf Schulebene aus der SDL verglichen.

Methode

Aus methodischer Sicht stellt sich zunächst die Frage, wie anderssprachige Schülerinnen und Schüler definiert werden. In der SDL des Schuljahres 2014/15 betrug der Anteil anderssprachiger 15-Jähriger 25.85 Prozent. Trotz dieses an der PISA-Schätzung (25.96 Prozent; 95%-Vertrauensintervall: 23.64 bis 28.27 Prozent; Anteil fehlender Antworten: 0.57 Prozent) äusserst nahen Anteils muss festgehalten werden, dass Anderssprachigkeit in den beiden Erhebungen nicht identisch definiert wurde. Während mit Hilfe der SDL die Erstsprache erhoben wird, wird im Rahmen des bei PISA eingesetzten Fragebogens nach der «normalerweise zu Hause gesprochenen Sprache» gefragt. Bei beiden Erhebungen werden Jugendliche als anderssprachig eingestuft, wenn die Erstsprache bzw. die zu Hause gesprochene Sprache nicht mit der Unterrichtssprache übereinstimmt.

Aufgrund dieser unterschiedlichen Operationalisierung von Anderssprachigkeit und der deshalb eingeschränkten Vergleichbarkeit zwischen SDL und PISA wird hier ein Vergleichsansatz präsentiert, der ausschliesslich auf Zahlen der SDL beruht, jedoch die Gewichte der bei PISA 2015 gezogenen Schulen mitberücksichtigt. Um zu überprüfen, inwieweit tatsächlich Schulen mit relativ hohen Anteilen anderssprachiger 15-Jähriger in der Stichprobe PISA 2015 übervertreten sind, wurden die bei PISA 2015 teilnehmenden Schulen mit der SDL verknüpft. Anschliessend wurde der Anteil Schülerinnen und Schüler in der Population, deren Erstsprache nicht der Schulsprache entspricht, ausschliesslich auf Basis der bei PISA 2015 teilnehmenden Schulen geschätzt. Dazu wurden die im Rahmen von PISA 2015 errechneten Schulgewichte sowie Replikationsgewichte auf Schulebene herangezogen und die entsprechenden Anteile unter Berücksichtigung der Schülerumfänge pro Schule auf die Schülerpopulation hochgerechnet.

Ergebnisse

Der Anteil 15-Jähriger, deren Erstsprache nicht der Schulsprache entspricht, betrug im Schuljahr 2014/15 laut SDL 25.85 Prozent. Werden die Zahlen der SDL nur auf Basis der bei PISA 2015 teilnehmenden Schulen auf die Schülerpopulation hochgerechnet, beträgt die Schätzung 27.98 Prozent mit einem 95-Prozent-Vertrauensintervall zwischen 25.96 und 30.00 Prozent. Es kann also festgehalten werden, dass mit der bei PISA 2015 gezogenen Schulstichprobe anderssprachige – im Sinne der Erstsprache – Schülerinnen und Schüler um ungefähr zwei Prozentpunkte übertreten sind. Das Vertrauensintervall der Schätzung liegt knapp über dem Populationsanteil. Zur Schätzung des Effekts einer solchen Differenz auf die Leistungsmittelwerte wurde der Anteil anderssprachiger Schülerinnen und Schüler im PISA-Datensatz um zwei Prozentpunkte reduziert, indem sämtliche Schülergewichte mithilfe eines Korrekturfaktors¹⁴ angepasst wurden. Dabei hat sich gezeigt, dass ein um zwei Prozentpunkte niedrigerer Anteil anderssprachiger Jugendlicher in einer Zunahme von 1.2 bis 1.5 Punkten auf der PISA-Skala in den drei untersuchten Domänen resultieren würde.

Diskussion

Die Güte einer Stichprobe lässt sich in Schulleistungsstudien daran messen, inwieweit sich mit der gezogenen Stichprobe ein Abbild der Population hinsichtlich der sie charakterisierenden und im Zusammenhang mit schulischen Leistungen stehenden Merkmale darstellen lässt (Brühwiler, Helmke, & Schrader, 2017; OECD, 2016a, 2016b; Rendtel, 1993; Rendtel & Pöttner, 1992). Im Zusammenhang mit der für PISA 2015 gezogenen Schweizer Stichprobe wurden in diesem Beitrag (1) die Häufigkeiten verschiedener Schulstufen und -programme, (2) die Anteile von Jugendlichen mit Migrationshintergrund anhand externer Daten sowie (3) die Quoten anderssprachiger Schülerinnen und Schüler validiert.

Wie aufgezeigt werden konnte, weisen erstens die mithilfe von PISA 2015 geschätzten Anteile von Schulstufen und Schulprogrammen eine hohe Ähnlichkeit mit offiziellen Statistiken zur Population auf. Die leichten Verschiebungen zwischen den beiden Sekundarstufen können zu grossen Teilen auf Differenzen zwischen der erwarteten und der tatsächlichen Anzahl unterrichteter Schülerinnen und Schüler in den gezogenen Schulen zurückgeführt werden. Die meisten der auf Sekundarstufe I unterrichteten Schulprogramme fielen in die durch PISA 2015 geschätzten Vertrauensintervalle. Die einzige Ausnahme bildete der Anteil in Sonderklassen unterrichteter 15-Jähriger. Es ist plausibel anzunehmen, dass der bei PISA zu niedrig geschätzte Anteil in Sonderklassen unterrichteter Jugendlicher auf Ausschlüsse auf Schülerebene zurückzuführen ist

(physische bzw. psychische Beeinträchtigungen oder zu geringe Kenntnisse der Testsprache). Der um ungefähr zwei Prozentpunkte höhere Anteil an Gymnasiasten der Sekundarstufe I in der PISA-Stichprobe lässt sich dadurch erklären, dass einzelne – in der SDL als Schulen ohne Selektion erfasste – Privatschulen ihre Schülerinnen und Schüler der Gruppe der Gymnasiasten zugeordnet haben.

Ein ebenso symmetrisches Bild zeigte sich auf der Sekundarstufe II: Einzig der Anteil der Schülerkategorie «Berufs- oder Fachmaturitäten» ist in der Population höher als das entsprechende mittels PISA 2015 geschätzte Vertrauensintervall. Die leicht verschobenen Werte bei den Fachmittelschulen, Berufs- oder Fachmaturitäten sowie bei den gymnasialen Maturitätsschulen könnten auf bestimmte Formulierungen im Erhebungsinstrument zurückzuführen sein. Da die Antwortoption der «Maturitätsschulen» lediglich mit einer Klammerbemerkung «für die gymnasiale Maturität» ergänzt war, ist nicht auszuschliessen, dass beispielsweise Berufsmaturitätsschulen die Option «Maturitätsschulen» angewählt haben. Generell kann jedoch festgehalten werden, dass die zahlreichen Schulprogramme der Schweiz adäquat in der PISA-Stichprobe vertreten sind.

Zweitens wurde deutlich, dass die Anteile von Schülerinnen und Schülern mit Migrationshintergrund bei PISA 2015 mit den Schätzungen der Schweizerischen Strukturerhebung grösstenteils übereinstimmen: Beide Quellen schätzten den Anteil von 15-Jährigen mit Migrationshintergrund auf ca. 30 Prozent. Leichte Differenzen fielen bei der Unterscheidung nach Migrationsgeneration auf: Während die Gruppe der zweiten Generation bei PISA leicht grösser ausfiel, wurde der Anteil Schülerinnen und Schüler mit Migrationshintergrund erster Generation bei PISA etwas niedriger geschätzt als mit den Daten der Strukturerhebung. Letzteres könnte damit zusammenhängen, dass bei PISA ein Teil der Lernenden mit schlechten Kenntnissen in der Testsprache von den verantwortlichen Lehrpersonen von der PISA-Teilnahme befreit wurde. Trotz der sehr ähnlichen Anteilsschätzungen sei aber darauf hingewiesen, dass die Vergleiche mit Vorsicht zu interpretieren sind. Während sich die Strukturerhebung auf die gesamte, ständige Wohnbevölkerung in Privathaushalten bezieht, besteht die PISA-Population aus Jugendlichen, die in einer Regelschule unterrichtet und gleichzeitig nicht von der Erhebung ausgeschlossen wurden. Es ist also anzunehmen, dass die beiden verglichenen Populationen nicht völlig deckungsgleich sind. Zudem unterscheidet sich der Zeitpunkt, an welchem die Jugendlichen mit Jahrgang 1999 zu ihrem Migrationshintergrund befragt wurden. Es dürfte sich bei beiden Quellen jedoch grösstenteils um dieselben Jugendlichen handeln (Jahrgang 1999).

Drittens kam zum Vorschein, dass mit den im Rahmen von PISA 2015 gezogenen Schulen der Anteil von anderssprachigen Schülerinnen und Schülern – im Sinne der Erstsprache – um ungefähr zwei Prozentpunkte überschätzt wurde. Wird berücksichtigt, dass Lernende mit schlechten Kenntnissen in der Testsprache teilweise von der Erhebung ausgeschlossen wurden, könnte die Überschätzung noch leicht stärker sein. Diese Differenz dürfte die nationalen

Leistungsmittelwerte trotzdem nur in einem nicht relevanten Ausmass beeinflussen. Die Güte der PISA-Schätzungen zu der zu Hause gesprochenen Sprache kann mangels offizieller Statistiken nicht abschliessend beurteilt werden. Hinzu kommt, dass die Frage nach der zu Hause gesprochenen Sprache – von 15-Jährigen in der Schweiz – häufig nicht eindeutig beantwortet werden kann.

Weiter gilt es die Umstellung auf computerbasiertes Testen in Erinnerung zu rufen. Dabei steht fest, dass die Aufgaben aller erhobenen Leistungsdomänen im Mittel schwieriger sind, wenn diese am Computer anstatt auf Papier bearbeitet werden (OECD, 2017; Robitzsch et al., 2017). Die genaue Ausprägung dieses Moduseffekts scheint jedoch von diversen Faktoren abhängig zu sein (Jerrim, Micklewright, Heine, Sälzer, & McKeown, 2018). Deshalb gestaltet es sich zum aktuellen Zeitpunkt äusserst schwierig, die eingangs erwähnten Leistungsdifferenzen zwischen PISA 2012 und PISA 2015 in der Schweiz zu genauen Teilen dem Moduseffekt, den weiteren methodischen Änderungen der OECD (z.B. Skalierungsmodell), der Stichprobenzusammensetzung und allfälligen tatsächlichen Leistungseinbussen zuzuordnen.

Auf die Qualität der Stichprobe hat der Erhebungsmodus grundsätzlich keinen Einfluss. Es ist jedoch theoretisch denkbar, dass die im vorliegenden Text zur Überprüfung der Stichprobenqualität herangezogenen Populationsmerkmale in gewisser Weise vom Erhebungsmodus abhängig sind. So könnte sich vor allem das Antwortverhalten mehrsprachiger Schülerinnen und Schüler bei Fragen zu den im familiären Umfeld gesprochenen Sprachen zwischen den beiden Fragebogenmodi – hauptsächlich aufgrund der erzwungenen Einfachantwort im computerbasierten Modus – unterscheiden. Die geschätzten Anteile zum Migrationsstatus dürften hingegen nicht von einem derartigen Moduseffekt betroffen sein (Konsortium PISA.ch, 2018). Über weitere Ursachen für die Differenzen in der Stichprobenzusammensetzung zwischen PISA 2012 und PISA 2015 kann an dieser Stelle nur spekuliert werden. So könnten nebst den weiter oben beschriebenen zufälligen (unsystematischen) Stichprobenfehler auch systematische Fehler (Bias) eine Rolle gespielt haben. Zusammenhänge zwischen Ziehungsmechanismus und individuellen Merkmalen treten bei PISA aufgrund des eingesetzten Stichprobenziehungsverfahrens zwar zufällig auf. Stichprobenfehler aufgrund schwer kontrollierbarer Zusammenhänge zwischen Antwortverweigerungen, Absenzen oder Ausschlüssen und individuellen Merkmalen können jedoch nicht ausgeschlossen werden.

Aufgrund der nationalen, praktisch-bildungspolitischen Orientierung dieses Beitrags wird hier der internationale Forschungsstand weder zu inhaltlichen Aspekten noch zur Qualitätsbeurteilung von Stichproben oder zur Konzeption und Durchführung von Large Scale Assessments aufgezeigt. Ebenso wenig wird die Relevanz dieser Aspekte für das schweizerische Bildungswesen geprüft. Stattdessen zeigt dieser Beitrag auf, dass Differenzen der nationalen Leistungsmittelwerte der Schweiz zwischen PISA 2012 und PISA 2015 nur zu einem sehr geringen Ausmass auf die Stichprobenzusammensetzung zurückgeführt werden können.

Darüber hinaus wurde für die Schülermerkmale Schulprogramm, Migrationsstatus sowie Anderssprachigkeit eine hohe Kongruenz zwischen Häufigkeiten in der Population und in der Stichprobe PISA 2015 berichtet. Daraus folgt, dass die mithilfe von PISA 2015 gewonnenen Daten valide Aussagen über 15-jährige Schülerinnen und Schüler, das Bildungssystem der Schweiz und internationale Systemvergleiche ermöglichen.

Nachdem die berichteten Analysen die Gültigkeit der aktuellsten PISA-Stichprobe stützen, stellt sich die Frage nach den Gründen für die unterschiedlich geschätzten Anteile anderssprachiger bzw. immigrierter 15-Jähriger zwischen den Erhebungen 2012 und 2015. Gerade hinsichtlich einer allfälligen Optimierung des Stichprobenziehungsverfahrens für zukünftige PISA-Erhebungen ist dieser Aspekt von zentraler Bedeutung. Weiter lassen sich in Kürze die Daten aus der Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in der Schweiz aus dem Jahre 2016 sowie diejenigen aus PISA 2018 zur erweiterten externen Validierung verschiedener hier herausgearbeiteter Kennwerte heranziehen. Dies wird insbesondere hinsichtlich der Prüfung des Anteils anderssprachiger Schülerinnen und Schüler relevant, da diese Information in diesen Untersuchungen auf identische Art und Weise erfasst wurde. Hierin liegt auch die zentrale Bedeutung des Bildungsmonitorings: Eine regelmässige und zeitnahe sowie über die Zyklen hinweg möglichst parallele Erfassung relevanter Merkmale des Bildungssystems, der Schulen, der Schülerschaft sowie deren Leistungen ist zentral für eine frühzeitige Erkennung allfälliger Veränderungen.

Anmerkungen

- ¹ Anderssprachig oder Anderssprachigkeit bedeuten in diesem Kontext, dass die Erstsprache resp. die normalerweise zu Hause gesprochene Sprache (je nach Operationalisierung) nicht der Schulsprache entspricht.
- ² Bei der Autorenschaft dieses Beitrags handelt es sich aufgrund der Zugehörigkeit zum Konsortium PISA.ch um eine vertrauenswürdige Quelle in Bezug auf das Stichprobenziehungsverfahren und die Durchführung von PISA 2015. Das international standardisierte methodische Vorgehen bei PISA 2015 wird im technischen Bericht der OECD (OECD, 2017) beschrieben. Diverse für die Schweiz spezifische Angaben zum Stichprobenziehungsverfahren oder zur Datenerhebung wurden bisher jedoch nicht dokumentiert. Dementsprechend stellt der vorliegende Beitrag auch eine erstmalige Beschreibung der in der Schweiz verwendeten Stichprobenmethodik dar.
- ³ Dem sechs Wochen langen Testzeitfenster entsprechend, in welchem die Erhebungen an den Schulen durchgeführt werden müssen, sind die Schülerinnen und Schüler zum Zeitpunkt der Erhebung zwischen 15 Jahren und 3 Monaten und 16 Jahren und 2 Monaten alt (OECD, 2017).
- ⁴ Vgl. als Grundlage die Schweizerische Lernendenstatistik des BFS: <https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/erhebungen/sdl.html> (29.11.2018).
- ⁵ Das Ziehen zusätzlicher kantonalen Stichproben führte zu einer hohen Varianz der Stichprobengewichte. Der daraus resultierende, vergleichsweise hohe Designeffekt (vgl. Le, Brick, & Kalton, 2002) hob die Vorteile eines grösseren Stichprobenumfangs wieder auf.
- ⁶ Die Jahre 2000 und 2003 wurden hier nicht berücksichtigt. In diesen beiden Jahren war das Erhebungsdesign komplex (z.B. Zusatzstichproben in einigen wenigen Kantonen).

Das Einbeziehen dieser Angaben wäre im Rahmen der vorliegenden Darstellung, welche die Bedeutung des Stichprobenumfangs für die Zuverlässigkeit der Schätzung aufzeigen will, nicht förderlich.

- 7 Die Sortierung der Stichprobeneinheiten hilft bei solchen systematischen Ziehungsverfahren ebenfalls, die Häufigkeiten der Merkmale, die zur Sortierung herangezogen werden, besser entsprechend der Population abzubilden als bei reinen Zufallsstichproben. Der Ausdruck «systematisch» bedeutet in diesem Zusammenhang, dass ein auf einer Zufallszahl beruhendes Ziehungsintervall definiert wird, mit welchem die sortierten Listen «durchgezählt» werden. Die entsprechenden «Treffer» gelten dann als gezogene Einheiten. Da das Ziehungsverfahren die Abfolge solcher Listen genau festlegt, werden die gezogenen Schulen sowie Schülerinnen und Schüler (die Systematik gilt für beide Ziehungsebenen) einzig durch eine Zufallszahl bestimmt.
- 8 In gezogenen Schulen mit 35 oder weniger 15-jährigen Schülerinnen und Schülern wurden Vollerhebungen durchgeführt. Diese Klumpengrösse von 35 Individuen pro Schule wird von der OECD vorgegeben und gilt als guter Kompromiss zwischen Messgenauigkeit und organisatorischem Aufwand. Eine Aufteilung der Stichprobe auf mehr Schulen und eine entsprechende Reduktion der Klumpengrösse würde zwar zu präziseren Ergebnissen führen, sie wäre aber mit einem deutlich höheren organisatorischen sowie finanziellen Aufwand verbunden.
- 9 Hier handelt es sich lediglich auf theoretischer Ebene um eine EPSEM-Stichprobe («equal probability selection method»). Da die Schulgrössen auf Basis von Daten vergangener Schuljahre geschätzt werden müssen, passen die Schulauswahlwahrscheinlichkeiten nicht immer genau zur Anzahl unterrichteter Schülerinnen und Schüler. Dies führt dazu, dass auch Schülerinnen und Schüler verschiedener Schulen desselben Stratoms nur selten die exakt identischen Auswahlwahrscheinlichkeiten aufweisen.
- 10 Der Vergleich wird leicht dadurch eingeschränkt, dass die Zahlen der SDL sämtliche an Regelschulen unterrichteten Schülerinnen und Schüler enthalten. Die teilweise im Rahmen von PISA durch Schulpersonal ausgeschlossenen Schülerinnen und Schüler werden dort also mitberücksichtigt.
- 11 Hierbei handelt es sich um eine offiziell durch die OECD erhobene Variable, die im öffentlichen, internationalen Datensatz verfügbar ist (TFStudyProg; siehe PISA Database unter <http://www.oecd.org/pisa/data/2015database/>). Die Anforderungsniveaus bzw. kantonalen Schulprogramme wurden in eigener Regie durch die regionalen Durchführungszentren des PISA-Konsortiums Schweiz erhoben.
- 12 Sonderklassen werden innerhalb der Regelschule geführt und umfassen Schülerinnen und Schüler, die im Regelunterricht als gefährdet betrachtet werden oder nicht adäquat gefördert werden können (EDK, 2007; EDK/IDES, 2007). Sonderklassen wurden im Gegensatz zu Sonderschulen in die Stichprobe miteinbezogen.
- 13 Die Anzahl fehlender Angaben wurde verringert, indem Schülerinnen und Schüler, die beim Geburtsort eines Elternteiles «Schweiz» angegeben und beim zweiten Elternteil keine Antwort gegeben haben, der Gruppe «ohne Migrationshintergrund» zugeordnet wurden.
- 14 Dieser Korrekturfaktor wurde getrennt für anderssprachige und nicht-anderssprachige Schülerinnen und Schüler berechnet, indem der entsprechende Populationsanteil gemäss SDL mit demjenigen in der PISA-Stichprobe dividiert wurde. Die angepassten Gewichte resultierten dementsprechend in derselben Summe, jedoch in einem mit der SDL übereinstimmenden Anteil anderssprachiger Schülerinnen und Schüler. Es wurden keine Stratifikationsvariablen bei dieser Nachgewichtung berücksichtigt.

Literatur

- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. überarb. u. akt. Aufl.). Berlin: Springer.
- Brühwiler, C., & Helmke, A. (2018). Determinanten der Schulleistung. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Hrsg.),
- Brühwiler, C., Helmke, A., & Schrader, F.-W. (2017). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 291-314). Wiesbaden: Springer VS.
- Bundesamt für Statistik. (2017). *Statistik der Lernenden (Schüler/innen und Studierende). Technisches Handbuch der Erhebung 2017/18*. Zugriff am 6.2.2018. Verfügbar unter <https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/publikationen.assetdetail.2280923.html>
- Dachverband Lehrerinnen und Lehrer Schweiz. (2016). *PISA 2015: Viele Fragezeichen und keine neuen Erkenntnisse*. Verfügbar unter https://www.lch.ch/fileadmin/files/documents/Medienmitteilungen/161206_Medienmitteilung_LCH_PISA_2015.pdf
- EDK. (2007). *Einheitliche Terminologie für den Bereich der Sonderpädagogik*. Verfügbar unter https://www.edudoc.ch/static/web/arbeiten/sonderpaed/terminologie_d.pdf
- EDK. (2009). *Jahresversammlung der EDK vom 29. und 30. Oktober 2009*. Verfügbar unter http://edudoc.ch/record/35600/files/education_3_d.pdf
- EDK. (2012). *Teilnahme der Schweiz an PISA 2015. Beschluss Plenarversammlung*. Verfügbar unter https://edudoc.ch/record/105012/files/PB_pisa_d.pdf
- EDK. (2016). *PISA 2015: Neustart mit Fragezeichen*. Verfügbar unter <http://www.edk.ch/dyn/30196.php>
- EDK/IDES. (2007). *Sonderpädagogik. Kapitel 10 des Schweizer Beitrags für die Datenbank «Eurybase - The database on education systems in europe»*. Verfügbar unter http://www.edk.ch/dyn/bin/12961-13439-1-eurydice_10d.pdf
- Forschungsgemeinschaft PISA Deutschschweiz/FL. (2005). *PISA 2003: Analysen für Deutschschweizer Kantone und das Fürstentum Liechtenstein. Detaillierte Ergebnisse und methodisches Vorgehen*. Zürich: Kantonale Drucksachen- und Materialzentrale.
- Jerrim, J., Micklewright, J., Heine, J.-H., Sälzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education*, 44, 476-493.
- Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, 6(3), 223-239.
- Konsortium PISA.ch. (2015). *PISA 2015: Fragebogen für Schülerinnen und Schüler. Instrument für die computerbasierte Haupterhebung PISA 2015*. Verfügbar unter <https://www.education.ch/sites/default/files/uploads/2017/10/schuker.pdf>
- Konsortium PISA.ch. (2018). *PISA 2015: Schülerinnen und Schüler der Schweiz im internationalen Vergleich*. Bern und Genf: SBFI/EDK und Konsortium PISA.ch.
- Kruskal, W., & Mosteller, F. (1979a). Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review / Revue Internationale de Statistique*, 47(2), 111-127.
- Kruskal, W., & Mosteller, F. (1979b). Representative sampling, III: The current statistical literature. *International Statistical Review / Revue Internationale de Statistique*, 47(3), 245-265.
- Kruskal, W., & Mosteller, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895-1939. *International Statistical Review / Revue Internationale de Statistique*, 48(2), 169-195.
- Le, T., Brick, M., & Kalton, G. (2002). Decomposing design effects. In *JSM Proceedings, Survey Research Methods Section* (pp. 2007-2012). Alexandria, VA: American Statistical Association.
- Meinck, S. (2015). Computing sampling weights in large-scale assessments in education. *Survey insights: Methods from the field, weighting: Practical issues and 'how to' approach*.

- Retrieved from <http://surveyinsights.org/?p=5353>
- Neue Zürcher Zeitung. (2016, 6. Dezember). *Bestnoten für Schweizer Schüler im Fach Mathematik*. Verfügbar unter <https://www.nzz.ch/schweiz/pisa-studie-2015-schweizer-15-jaeh-rige-rechnen-europaweit-am-besten-ld.132847>
- OECD. (2009). *PISA data analysis manual*. Paris: OECD Publishing.
- OECD. (2014). *Student questionnaire for PISA 2015. Computer based version. Main survey version (CY6_QST_MS_STQ_CBA_Final)*. Retrieved from http://vs-web-fs-1.oecd.org/pisa/PISA2015-ENG_Quest.zip
- OECD. (2016a). *PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung*. Bielefeld: W. Bertelsmann Verlag.
- OECD. (2016b). *PISA 2015 results (volume II): Policies and practices for successful schools*. Paris: OECD Publishing.
- OECD. (2017). *PISA 2015 technical report*. Paris: OECD Publishing.
- Ostschweiz am Sonntag. (2016, 11. Dezember). «Pisa ist nationalen Tests um Jahre voraus» (S. 7.).
- Rendtel, U. (1993). Über die Repräsentativität von Panelstichproben: Eine Analyse der feldbedingten Ausfälle im Sozio-Oekonomischen Panel (SOEP). DIW Diskussionspapier, Nr. 70.
- Rendtel, U., & Pöttner, U. (1992). Über Sinn und Unsinn von Repräsentationsstudien. DIW Diskussionspapier, Nr. 61.
- Robitzsch, A., Lüdtke, O., Köller, O., Kröhne, U., Goldhammer, F., & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schulleistungstudien. Eine Skalierung der deutschen PISA-Daten. *Diagnostica*, 63, 148–165.
- Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, 1(4), 381–397.
- Rust, K. (2014). Sampling, weighting, and variance estimation in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 117–153). Boca Raton, FL: CRC Press.
- Rust, K., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3), 283–310.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151.
- Rutkowski, L., von Davier, M., & Rutkowski, D. (Eds.). (2014). *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: CRC Press.
- Ryan, K. E., & Shepard, L. A. (2008). *The future of test-based educational accountability*. New York, NY: Routledge.
- Schnell, R., Hill, P. B., & Esser, E. (1999). *Methoden der empirischen Sozialforschung* (6. Aufl.). München: Oldenbourg.
- Tagblatt. (2016, 6. Dezember). *PISA-Studie ärgert Schweizer Verantwortliche*. Verfügbar unter <http://www.tagblatt.ch/nachrichten/schweiz/PISA-Studie-aergert-Schweizer-Verantwortliche;art253650,4846576>
- von der Lippe, P., & Kladroba, A. (2002). Repräsentativität von Stichproben. *Marketing ZFP – Journal of Research and Management*, 24, 139–144.
- Wolter, K. M. (2007). *Introduction to variance estimation*. New York, NY: Springer.

Schlagworte: Internationale Schulleistungstudie, PISA, Schweiz, Zufallsstichprobe, Migrationshintergrund, Schulprogramm

L'échantillon suisse de PISA 2015 – Une validation externe des caractéristiques principales

Résumé

Les enquêtes internationales sur les compétences des élèves n'évaluent pas les performances individuelles, mais celles des populations visées. Pour des raisons pratiques, on recourt au tirage d'échantillons. Afin de parvenir à des conclusions valides sur les populations cibles, on recourt à des procédures appropriées de tirage et de pondération de l'échantillon. En prenant PISA 2015 pour exemple, le texte illustre les critères de qualité d'un échantillon aléatoire scientifique. Pour ce faire, la procédure d'échantillonnage utilisée en Suisse est décrite. De plus, on vérifie dans quelle mesure l'échantillon suisse PISA 2015 représente correctement la population cible, notamment en termes de programme scolaire, de statut migratoire et des autres langues parlées. Les analyses confirment la validité de l'échantillon suisse PISA 2015.

Mots-clés: Enquête internationale sur les compétences des élèves, PISA, Suisse, échantillon aléatoire, statut migratoire, programme scolaire

Il campione svizzero PISA 2015 – Una validazione esterna delle caratteristiche principali

Riassunto

Gli studi internazionali sulle prestazioni scolastiche non si concentrano sulle prestazioni individuali, ma su quelle a livello di popolazioni. Per motivi pratici, si ricorre all'utilizzo di campioni di allievi. Per trarre delle conclusioni valide sulle popolazioni, è necessario adottare procedure adeguate di estrazione e di ponderazione di questi campioni. Utilizzando l'indagine PISA 2015 come esempio, il testo illustra i criteri di qualità dei campionamenti scientifici casuali. In questa occasione viene descritta la procedura di campionamento utilizzata in Svizzera. Inoltre, si valuta in che misura il campione svizzero di PISA 2015 descriva adeguatamente le caratteristiche della popolazione rispetto ai programmi scolastici, allo statuto migratorio e alle altre lingue parlate. Le analisi confermano la validità del campione svizzero di PISA 2015.

Parole chiave: Studio internazionale sulle prestazioni scolastiche, PISA, Svizzera, campione casuale, statuto migratorio, programma scolastico

On the Swiss PISA 2015 sample - An external validation of core sample characteristics

Summary

International educational assessments do not focus on individual performance assessments, but rather make inferences about populations. For practical reasons, student samples are drawn. In order to allow valid conclusions about populations on the basis of samples, methodologically appropriate drawing and weighting procedures are necessary. Using PISA 2015 as an example, the text at hand illustrates quality characteristics of scientific random samples. Besides, the sampling procedure used in Switzerland is described. In addition, it will be examined to what extent the Swiss sample for PISA 2015 adequately depicts population characteristics, such as school programmes or the migration background and other spoken languages. The analyses support the validity of the Swiss sample for PISA 2015.

Keywords: International educational assessments, PISA, Switzerland, random sample, migrant background, school program

Martin Verner, Dr. phil., Studium der Psychologie (Persönlichkeits-, Differentielle Psychologie und Diagnostik) und Geschichte sowie Promotion an der Universität Bern. Seit 2014 am Institut für Bildungsevaluation in Zürich als wissenschaftlicher Mitarbeiter tätig. Arbeitsschwerpunkte: Skalierung von schulischen Leistungsdaten, Stichproben- und Datenmanagement bei kantonalen/nationalen Projekten (u.a. PISA, ÜGK). Institut für Bildungsevaluation, assoziiertes Institut der Universität Zürich, Wilfriedstrasse 15, CH-8032 Zürich. E-Mail: martin.verner@ibe.uzh.ch

Andrea B. Erzinger, Dr. phil, Lizentiat in Pädagogik, Privatrecht und Politikwissenschaft; Promotion in Erziehungswissenschaft an der Universität Zürich. Forschungsschwerpunkte: Kompetenzentwicklung im Kindes- und Jugendalter, Interaktion Schule und Familie, Intergenerationale Beziehungen im Lebenslauf. Nationale Projektmanagerin PISA. Tätig als Leiterin des Interfaculty Centre for Educational Research (ICER) an der Universität Bern, Fabrikstrasse 8, CH-3012 Bern. E-Mail: andrea.erzinger@icer.unibe.ch

Ursina Fässler, M.Sc., Studium in Psychologie und Erziehungswissenschaften-Sonderpädagogik an der Universität Zürich sowie Master of Advanced Studies in Kinder- und Jugendpsychologie an der Universität Basel. 2013 bis 2016 tätig als Schulpsychologin. Seit 2017 wissenschaftliche Mitarbeiterin in den Projekten PISA und ÜGK. Pädagogische Hochschule St.Gallen, Notkerstrasse 27, CH-9000 St.Gallen. E-Mail: ursina.faessler@phsg.ch